Language Fluency and Earnings

Dustmann, C.; van Soest, A.H.O.

Publication date:
1998

Contents

Language Fluency and Earnings: Estimation with Misclassified Language Indicators

Christian Dustmann[†] Arthur van Soest[‡]

November 3, 1998

JEL codes: C23, J61

Abstract

We use panel data from the German Socio Economic Panel to estimate the determinants of language fluency of immigrants, and its impact on earnings. Self reported measures of language proficiency contain substantial reporting errors. We specify a panel data model which takes explicitly account of misclassification. We extend the existing literature on misclassification of categorical dependent variables by distinguishing between time persistent and time varying misclassification errors, using panel data. The repeated information on language fluency allows us also to distinguish between cohort effects and exposure effects. We then add a wage equation to the model and estimate it jointly with the speaking fluency equation. In this way, we take account of the two problems that may bias OLS estimates: misclassification errors and correlated unobserved individual heterogeneity in wages and speaking fluency. We find that both have important consequences for the estimated effect of speaking fluency on earnings.

---

*We are grateful to Costas Meghir and Marcel Das for comments on earlier drafts of this paper.

[†]University College London, Department of Economics, and Institute for Fiscal Studies, London. e-mail: c.dustmann@ucl.ac.uk

[‡]Tilburg University, P.O. Box 90153, 5000 LE Tilburg, Netherlands. e-mail: avas@kub.nl

1 Introduction

An important component of host country specific human capital of migrant workers are linguistic skills. Many recent studies have investigated the determinants of language proficiency, and have analyzed the effect of dominant language fluency on labor market performance (see, for instance, Carliner (1981), McManus et al. (1983), Grenier (1984), Kossoudji (1988), Rivera-Batiz (1990), Chiswick (1991), Dustmann (1994), and Chiswick and Miller (1995)). Most of these studies conclude that language proficiency is higher the higher the level of education, and the lower the age at arrival, and that it improves substantially with the time spent in the host country. Furthermore, language fluency is positively related to earnings.

All these studies draw on cross–sectional data, and most use self–reported language ability as a measure for language proficiency. This variable suffers from misclassification error. This may lead to biased estimates when estimating traditional nonlinear models for discrete dependent variables. Hausman et al. (1997) demonstrate that even small probabilities of misclassification may lead to a large bias in parameter estimates in a probit model.

With self reported language fluency, like with other categorical variables which are based on subjective evaluations, there are two sources of misclassification error. First, random misclassification, independent over time. This is the type of error researchers typically have in mind when specifying models which explicitly take account of misclassification. For example, Hausman et al. (1997) model job changes. As a source of misclassification error they consider recall error or misunderstanding of survey questions. Second, like other variables which are based upon subjective standards, self

reported language fluency is likely to suffer, in addition, from a time persistent misclassification error, which reflects an individual specific over- or underevaluation of the true ability. With cross–section data, these two sources of misclassification can not be distinguished. With panel data, the two errors are in principle identified.

We use data from the German Socio-Economic Panel, covering the time period 1984 – 1993, with information on language fluency in seven waves. We estimate a model which takes explicitly account of misclassification errors. We draw here on work by Lee and Porter (1984) and Hausman et al. (1997), and generalize their approach to an ordered probit random effects panel data model. We distinguish between time varying and time persistent measurement errors, which are identified by the additional variation within individuals when using panel data. We use flexible specifications of the distributions of the random individual effects, following Heckman and Singer (1984). By explicitly modeling misclassification probabilities, we investigate the potential bias of the effect of covariates on language fluency.

Most studies cited above use OLS to estimate the effect of language on earnings. The common finding is that (self–reported assessment of) language proficiency has a significant positive impact on earnings. One source of bias has been addressed in this literature: the positive correlation between unobserved heterogeneity in earnings and speaking fluency equations, leading to upward biased estimates. Borjas (1994) argues that for this reason the effect of language fluency on earnings may be much lower than OLS estimates indicate. Chiswick and Miller (1995) use IV (instrumental variables) estimation to account for these problems. They compare OLS and IV estimates using data for different countries. Their results are rather unstable and imprecise, but in most of their estimations, the sign of the OLS bias of the language ability variable

points in the opposite direction of what unobserved heterogeneity would suggest. One explanation is that forgone earnings of individuals who engage in language education increase with their unobserved ability (see Willis and Rosen (1979)). In this case, unobserved heterogeneity may lead to downward biased estimates. Another explanation is measurement error. Both types of misclassification error in the language indicator lead to downward biased OLS estimates.

We analyze the effect of language on earnings, distinguishing between the potential bias induced by measurement error and correlated heterogeneity. We extend our panel data model for speaking fluency by adding a wage equation. The wage equation allows for unobserved individual heterogeneity which can be correlated with unobserved individual effects in the equation for speaking fluency. We use mass point distributions for both models, thus avoiding distributional assumptions. Moreover, since allowing for misclassification errors in speaking fluency removes measurement error of the fluency variable, the wage equation includes the true speaking fluency indicator as explanatory variable. Thus our model accounts for both types of bias. By estimating the model without allowing for misclassification errors and without allowing for correlated unobserved heterogeneity terms, we are able to analyze the consequences of both types of bias separately.

We also estimate specifications which allow, in addition, for correlation of the residual error terms in language and fluency equation. To identify the wage equation nonparametrically in this most general model, we need instruments in the speaking fluency which do not directly affect wages. Our data contains more background variables than usually available in these type of studies. This should bring us in a privileged position compared to Chiswick and Miller (1995), whose instruments (family composition

variables and a regional concentration of immigrants index) seem not very powerful.

A most robust finding in the literature on the determinants of language fluency is that the duration in the destination country has a strong positive effect on language fluency. The usual interpretation is that duration is an index of exposure, which steadily increases language fluency (see Chiswick and Miller (1995)). Besides the potential bias in this variable induced by a misclassified language index, results based on a single cross–section may be misleading, since cross–section data does not allow to distinguish between years of residence effects and cohort effects. If cohorts differ in the accumulation of language capital, the coefficient of the residence variable is biased. This is similar to the potential bias of assimilation effects in the literature on migrants' earnings adjustment (see Borjas (1985)). We show that neglecting cohort effects leads to biased estimates of the effect of years of residence on language fluency of migrants.

The paper is organized as follows. In section 2, we present the data. Section 3 presents the model for language ability. Section 4 discusses the wage equation and its estimates. Section 5 concludes.

2 Data

The data we use is drawn from the German Socio-Economic Panel (GSOEP), which started in 1984.[1] To our knowledge, the GSOEP is the only household panel which oversamples immigrants and provides therefore a sufficient database for statistical analyses of these minorities. In the first wave, it includes about 1500 households with a foreign born head. Foreign born individuals are asked a number of specific questions regarding

[1]See Wagner et al. (1993) for details on the GSOEP.

their economic behavior, as well as their economic and social integration. In the years 1984 - 1987, 1989, 1991, and 1993, questions are included regarding language fluency. Language information is not reported in the 1988, 1990, and 1992 surveys.

Speaking fluency is reported on a five point scale, with possible answers very bad (1), bad (2), intermediate (3), good (4), and very good (5). In our analysis, we consider males only. Our sample includes 1613 individuals in the first wave who provide information on self assessed language fluency. Due to missing information on explanatory variable, 83 of these could not be used in the analysis, leaving 1530 observations for the first wave. Due to attrition, the panel is unbalanced. From wave 1 to 2, we lose about 15 percent of our sample observations, in later waves, attrition is smaller. The numbers of observations used for the analysis are 1530 in 1984, 1299 in 1985, 1237 in 1986, 1210 in 1987, 1069 in 1989, 1024 in 1191, and 958 in 1993.

In Table 1, bivariate frequency distributions of self-reported speaking fluency in consecutive years are presented for the first four waves. The non-diagonal cells refer to changes in speaking fluency. There are many transitions from good to intermediate, from intermediate to bad or very bad, etc. Although some deterioration of speaking fluency is in principle possible, the large number of below-diagonal observations strongly suggests that the self-reported language ability measure suffers from misclassification errors which vary over time.

In table 2, we have summarized the changes in the speaking fluency variable (treated as a cardinal variable with values 1, 2, 3, 4 and 5) between two consecutive years, again for the first 4 waves. These numbers illustrate the magnitude of potential misclassification in this type of data. The distribution of changes is nearly symmetric, with similar numbers of individuals reporting deterioration and improvements between

Table 1: Cross-Tabulations, Language Fluency						
	1	**2**	**3**	**4**	**5**	Total
	vertical: 1984; horizontal: 1985					
1	**4**	9	4	0	1	18
2	7	**87**	68	19	1	182
3	3	78	**253**	137	14	485
4	1	21	108	**250**	59	439
5	0	2	17	67	**109**	195
Total	15	197	450	473	184	1319
	vertical: 1986; horizontal: 1987					
1	**4**	6	3	1	0	14
2	6	**99**	61	8	0	174
3	1	58	**235**	102	10	406
4	0	8	120	**259**	47	434
5	1	1	8	62	**95**	167
Total	1 2	172	427	432	152	1195
	vertical: 1988; horizontal: 1989					
1	**4**	6	1	2	0	13
2	3	**95**	61	8	1	168
3	0	50	**258**	104	8	420
4	1	8	93	**277**	42	421
5	0	0	9	57	**92**	158
Total	8	159	422	448	143	1180

1: very bad; **2**: bad; **3**: intermediate; **4**: good; **5**: very good.

years. Overall, about 57 percent of individuals do not report any changes, while 19 percent reports a deterioration by one category, and 2.2 percent by more than one category. A decrease of self-reported fluency therefore suggests that the self-reported measure is either too optimistic in the first year or too pessimistic in the second year. Assuming that deterioration of language fluency is not possible, and that the distribution of misclassification is symmetric, these numbers suggest that, on average, most of the within individual variation of language proficiency is due to misclassification.

To be precise: the total variance in the language indicator (on the cardinal 1 to 5 scale, all years) of 0.891 can be decomposed in a within individuals variance of 0.253, and a between individuals variance of 0.638.

If we assume that all reported deterioration is misclassification, that misclassification errors u_t are non negatively correlated over time, have a time independent variance, and the distribution of $u_t - u_{t-1}$ is symmetric around zero, then the variance of the measurement error satisfies $V(u_t) \geq P(y_t - y_{t-1} \leq -1)$, where y_t is observed speaking fluency.[2] According to Table 2, an estimate for this lower bound on the variance of the measurement error is then 0.213. Thus, under the assumption that deterioration is impossible, most of the within individuals variance and at least about one fourth of the total variance is explained by measurement error.

Additional to the potential misclassification errors revealed in the tables, some people may tend to persistently over- or underreport their language ability. Time persistent misclassification error does not show in our cross-tabulations. For example, a respondent who always reports "good" may indeed always have good proficiency,

[2]The assumptions and Chebyshev's Rule imply $V(u_t) \geq V(u_t) - Cov(u_t, u_{t-1}) = 0.5V(u_t - u_{t-1}) \geq 0.5P(|u_t - u_{t-1}| \geq 1) = P(u_t - u_{t-1} \leq -1) \geq P(y_t - y_{t-1} \leq -1) = 0.213$.

Table 2: Category changes, consecutive years									
	Deterioration								Improvement
Changes	**-4**	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**
Number of Obs.	1	5	75	709	2121	702	75	5	1
Percent	0.03	0.14	2.03	19.19	57.42	19.00	2.03	0.14	0.03
Notice: Numbers refer to the years 1984-1987.									

but may also always have "intermediate" proficiency and have a persistent tendency to overreport.

To model language fluency, we will use the set of standard regressors in these models.[3] Additional to the years since migration variable, we also include the year of entry. The latter picks up potential differences between different cohorts of migrants, i.e. between groups of migrants who came to Germany in different years. Conditional on this variable, the former measures the exposure effect, net of cohort effects. We also include age at entry and total years of education, and dummy variables indicating the immigrants' nationality (Turkish, Jugoslavian, Greek, Italian, or Spanish). In all these countries German is neither the dominant language, nor is it the first foreign language taught at school. It is therefore likely that the individuals in our sample spoke no or little German upon immigration.

In addition, we include some variables which are not accounted for in most other studies. These are several characteristics of the spouse (education level, age, year of entry) and the education level of the immigrant's father. The latter is drawn from the third wave of the panel, which contains information on several parental characteristics. Exact definitions and summary statistics of all the variables in our sample can be found

[3]Chiswick and Miller (1995) provide a systematic discussion of the determinants of language fluency.

in Table A1 in the appendix.

Our earnings variable is the natural logarithm of gross monthly earnings. In the earnings regressions we only include individuals who were in full–time employment during the month to which the earnings information refers.

3 A Panel Data Model for Language Ability

We observe speaking fluency on an ordinal scale with five categories. Because of the small number of observations in the extreme categories, we have combined levels 1 and 2 and levels 4 and 5, retaining categories "bad" ($y_{it} = 1$), "intermediate" ($y_{it} = 2$), and "good" ($y_{it} = 3$), where i is the individual and t the time period. As illustrated above, the raw data suggests that the language information is strongly affected by misclassification. We will introduce a panel data ordered response model which explicitly accounts for misclassification. The starting point is the random effects ordered probit model:

$$y_{it}^* = x_{it}'\beta + \alpha_i + \epsilon_{it}\,, \tag{1}$$

$$z_{it} = j \quad \text{if} \quad m_{j-1} < y_{it}^* < m_j, \quad j = 1, 2, 3\,, \tag{2}$$

$$\epsilon_{it} \text{ i.i.d. } N(0, \sigma_\epsilon^2)\,, \tag{3}$$

$$\alpha_i \text{ i.i.d. } N(0, \sigma_\alpha^2)\,, \tag{4}$$

$$\epsilon_{it}, \alpha_i \text{ and } x_{it} \text{ independent}\,. \tag{5}$$

Here x_{it} denotes the vector of explanatory variables, including a constant. Some of the x_{it} are constant over time (country of origin dummies, year of entry, age at entry), others vary over time (years of education, family composition and marital status, years

since migration), but not much or in a systematic way (years since migration, for example). α_i denotes the individual effect. Due to the lack of time variation in x_{it}, the data do not allow for estimating fixed effects models or models in which α_i is correlated with x_{it}. We do, however, relax the normality assumption on α_i. Following Heckman and Singer (1984), we replace (4) by the assumption that α_i follows a discrete distribution with K mass points:

$$P[\alpha_i = a_k] = p_k, \quad k = 1, \ldots, K\,. \tag{6}$$

The error term ϵ_{it} is idiosyncratic noise reflecting random variation in speaking fluency. In a model without explicit misclassification errors, this term will pick up measurement errors which are independent over time. If misclassification errors are explicitly incorporated, there is less scope for a meaningful interpretation of ϵ_{it}, and we would expect its impact (i.e. σ_ϵ) to be smaller. z_{it} denotes the speaking fluency category before misclassification, say the 'true' category. It is observed if there is no misclassification. By means of normalization, the category bounds are set to $m_0 = -\infty$, $m_1 = 0$, $m_2 = 10$, and $m_3 = \infty$.

To complete the model, we describe the relation between the observed category y_{it} and z_{it} in case of potential misclassification. Here we generalize models which explicitly allow for misclassification errors, such as Lee and Porter (1984), Douglas et al. (1995), and Hausman et al. (1997). These models do not distinguish between time-varying and time-persistent measurement error. Panel data allows to distinguish between both errors. As discussed above, categorical variables based on subjective evaluations are likely to suffer from time persistent misclassification error in addition. Our model uses the additional within individual variation to identify both sources of error. We use the

following assumptions, to allow for both types of misclassification in a parsimonious way:

- We distinguish four subpopulations: those who never misclassify ((0,0); fraction π_{00}), those who have a tendency to underreport but never overreport ((1,0); fraction π_{10}), those who have a tendency to overreport but never underreport ((0;1); fraction π_{01}), and those who can under– as well as overreport ((1;1); fraction $\pi_{11} = 1 - \pi_{00} - \pi_{01} - \pi_{10}$).

- The distributions of x_{it}, α_i, and ϵ_{it} are the same in each of the four subpopulations.

- Given the subpopulation and conditional on the true speaking fluencies (z_{it}), misclassification in different periods are mutually independent and independent of the x_{it}.

- The probabilities of underreporting for the subpopulations (1,0) and (1,1) do not depend on t and are given by $p_{21} = P[y_{it} = 1|z_{it} = 2]$, $p_{31} = P[y_{it} = 1|z_{it} = 3]$, and $p_{32} = P[y_{it} = 2|z_{it} = 3]$.

- The probabilities of overreporting for the subpopulations (0,1) and (1,1) do not depend on t and are given by $p_{12} = P[y_{it} = 2|z_{it} = 1]$, $p_{13} = P[y_{it} = 3|z_{it} = 1]$, and $p_{23} = P[y_{it} = 3|z_{it} = 2]$.

Thus, for example, the probability that an individual in subpopulation (1,0) with $z_{i1} = z_{i2} = 2$ gives answers $y_{i1} = 1$ and $y_{i2} = 2$, is given by $p_{21}(1 - p_{21})$. For someone in subpopulation (1,1), this probability is $p_{21}(1 - p_{21} - p_{23})$. For the other subpopulations, the probability is zero since these subpopulations never underreport.

Probabilities which are not conditional upon z_{it} can be written as weighted means of the probabilities given above, weighting with the probability distribution of the z_{it}. These probabilities still take the subpopulation as given, however. In practice, we do not observe in which subpopulation the respondents are. Likelihood contributions are therefore obtained by taking the weighted mean of the probabilities for each of the four subpopulations, using the subpopulation probabilities π_{00}, π_{01}, π_{10} and π_{11} as weights.

Obviously, this is not the only way to model misclassification explicitly. Compared to other ways, however, our model, has the advantage that it is relatively parsimonious (nine parameters: six p_{jk} $(j, k = 1, 2, 3, j \neq k)$ and π_{00}, π_{01} and π_{10}), but still nests the two extreme cases: misclassification which is independent over time, and misclassification which is completely time persistent. The former is obtained if $\pi_{11} = 1$; in this case, conditional upon true speaking fluencies z_{it}, events of misclassification are independent over time. The latter is obtained if, for example, $p_{21} = p_{31} = 1$. In this case, there is some fraction of people $(\pi_{10} + \pi_{11})$ who always report "bad" speaking fluency, whatever their real speaking fluency is.

In general, our model allows for any correlation between misclassification in two different time periods (conditional upon true speaking fluencies). For example, the probability that someone is reasonably fluent in both time periods, reports bad fluency twice, is given by $(\pi_{10} + \pi_{11})p_{21}^2$. The probability that this happens in one wave, is given by $(\pi_{10} + \pi_{11})p_{21}$. If $(\pi_{10} + \pi_{11}) = 1$, the two waves probability is the product of the two one wave probabilities, and the events in the two waves are independent. If $p_{21} = 1$, someone who once misreports reasonable as bad, will always do this as long as his true fluency is reasonable, so misclassification is time persistent. For other values of the parameters, any intermediate positive correlation structure can be obtained.

Negative correlations are not possible, and we do not consider them plausible in the current context.

Our specification is restrictive in the sense that the π_{qr} and the p_{jk} are not allowed to vary with x_{it} or t. They are treated as constant parameters. This assumption is also made in other (cross–section) studies with explicit misclassification errors, see Hausman et al. (1997), Lee and Porter (1984), and Douglas et al. (1995). The former two distinguish only two regimes, and thus work with two misclassification probabilities (the probability that the second regime is observed given that the first is true and vice versa), which are both treated as fixed parameters independent of everything else. Douglas et al. (1995) work with three (ordered) regimes but impose two misclassification probabilities to be zero, leaving them with four additional parameters to be estimated. Relaxing the assumption of fixed misclassification probabilities would require more from the data.

Our panel is unbalanced. We assume that whether information on individual i is available in wave t or not, is independent of $\{\epsilon_{it}, t = 1, \ldots, T\}$ and α_i. This implies that we do not allow for selection or attrition bias.

The model can be estimated by maximum likelihood. The assumptions given above imply that computing the likelihood contribution for each individual requires numerical integration in one dimension if the specification with normally distributed individual effects in (4) is used, as in the binary response case of Butler and Moffitt (1982). If the discrete distribution in (6) is used instead of (4), no numerical integration is required.

Results Speaking Fluency

We have estimated a large variety of specifications: with linear and non-linear cohort and years since migration effects, with and without explanatory variables referring to characteristics of the partner and education level of the father, with and without explicit unsystematic or time persistent misclassification, with normally distributed random effects and with Heckman–Singer type random effects. Four selected specifications are presented in Table 2. They all incorporate Heckman–Singer type random effects, based upon the discrete distribution given in (6), with four mass points. In terms of goodness of fit, the models with this type of random effects outperformed the models with normally distributed random effects. Both types of models lead to similar estimates of the other parameters in the model.

The first two specifications in Table 2 (Models 1 and 2) are standard panel data models with random effects, with no explicit misclassification errors. The first comes closest to the cross–section specifications in existing studies. Neither the cohort effect, nor partner's characteristics or the father's education level dummies are included. All these variables are included in the second model. The third and fourth specification use the same explanatory variables as model 2, but explicitly allow for misclassification. Model 3 allows for time independent classification errors, model 4 also allows for time persistent misclassification. Model 4 is the most general model, it nests the other three.

Most of the estimates of the slope coefficients β are robust across the four specifications, and also across the other specifications which are not presented. This also holds for their significance levels. Age at entry has a significantly negative impact, as in other studies. There are two explanations for this effect. First, learning a foreign language

Table 3: Estimation Results, Speaking Fluency								
	Model 1		**Model 2**		**Model 3**		**Model 4**	
	Coef	StdE	Coef	StdE	Coef	StdE	Coef	StdE
constant	-1.020	1.029	-11.536	3.805	-13.750	3.540	-12.386	4.021
year entry	–	–	0.106	0.051	0.112	0.046	0.112	0.051
age entry	-0.457	0.018	-0.448	0.023	-0.423	0.025	-0.477	0.032
d turkish	0.861	0.572	0.447	0.619	0.346	0.547	0.072	0.624
d jugos	5.017	0.613	4.597	0.637	4.345	0.585	4.608	0.698
d greek	2.517	0.636	1.776	0.630	1.748	0.556	1.919	0.654
d italian	-0.056	0.572	0.096	0.589	0.006	0.530	0.253	0.619
f educ l 2	–	–	0.629	0.455	0.731	0.405	0.579	0.472
f educ l 3	–	–	1.806	0.488	1.965	0.436	1.741	0.495
f educ l 4	–	–	4.026	0.971	4.481	0.902	4.231	1.023
f educ l 5	–	–	4.335	1.717	4.191	1.389	3.777	1.401
f educ l 6	–	–	0.540	0.577	0.817	0.515	1.068	0.626
yrs s migr	0.116	0.019	0.162	0.053	0.158	0.048	0.153	0.053
yrs educ	0.759	0.074	0.680	0.089	0.676	0.080	0.690	0.093
mar	-0.830	0.371	-0.519	0.394	-0.403	0.351	-0.402	0.379
mar, p G	4.697	0.778	4.324	0.891	3.892	0.847	4.256	0.900
p yrs edu			0.128	0.098	0.172	0.088	0.148	0.098
p yrs s migr			-0.032	0.042	-0.023	0.037	-0.029	0.043
p age			0.027	0.026	0.019	0.024	0.040	0.02 7
n children			-0.136	0.102	-0.133	0.095	-0.115	0.105
σ_ϵ	5.226	0.085	5.222	0.085	3.959	0.257	4.136	0.272
p12					0.132	0.035	0.315	0.067
p13					0.013	0.011	0.028	0.021
p21					0.039	0.009	0.444	0.072
p23					0.089	0.025	0.137	0.045
p31					0.001	0.001	0.004	0.008
p32					0	–	0	–
π_{00}							0.190	0.116
π_{01}							0.670	–
π_{10}							0.140	0.039
p_1	0.168	0.019	0.179	0.020	0.173	0.018	0.155	0.019
a_1	20.920	0.644	20.488	0.659	19.946	0.806	20.066	1.100
p_2	0.390	0.034	0.415	0.033	0.424	0.031	0.377	0.038
a_2	12.601	0.478	12.586	0.492	12.719	0.589	12.540	0.840
p_3	0.331	0.034	0.306	0.033	0.298	0.030	0.358	0.038
a_3	7.347	0.386	7.283	0.410	7.634	0.489	7.402	0.647
log lik	-5249.64		-5231.31		-5209.42		-5192.05	

becomes more difficult with age, leading to slower acquisition of language capital of those who immigrate later in life. Second, older migrants have a shorter pay off period on any country specific human capital, and this provides a disincentive effect.

The country dummies reflect distance in culture and language similarity. Country dummies could also reflect different degrees of self selection from different origin countries. They indicate that Jugoslavian immigrants are more fluent than the other groups, *ceteris paribus*. Greek immigrants are less fluent than Jugoslavians, but more fluent than the other three groups. Among these three (Turkish, Italian, and Spanish immigrants), differences are insignificant.

Years of education have a significant positive effect which is similar according to all specifications. The higher educated speak German more fluently than those with lower education level. This is also in line with the existing empirical evidence.

Separate dummy variables are included for men who are married with a foreign born partner, and who are married with a partner born in Germany. The reference group consists of unmarried men. The results indicate that men whose partner is German born are more fluent in German than the other two groups. Moreover, married men with a foreign born partner are less fluent than all others.[4]

Model 2 is a significant improvement of model 1. This is due to including the father's education level and to allowing for a cohort effect. On the other hand, the partner variables and the number of children in the family, are all insignificant, and a Wald test also indicates that they are jointly insignificant.[5]

[4]To make this interpretation possible in models 2, 3 and 4, the partner characteristics are set to their sample means for those without partner.

[5]This is not the case if the father's education level is excluded. In that case, both the partner's

As expected, the impact of the father's education level is positive. Children from families with higher educational background may be more likely to develop an interest for all those goods which are accessible with language. They may also grow up in a more open minded environment, reducing barriers to contacts to foreign cultures later in life. Furthermore, they may also be more likely to be exposed to a foreign language during their childhood.

The cohort effect is reflected by the year of entry into Germany. The results of models 2, 3 and 4 all point all in the same direction: later cohorts of immigrants speak German more fluently, conditional upon years since migration, age at entry, nationality, etc. Notice that this is conditional on country of origin dummies.[6] These effects may be explained by the specific type of immigration to Germany. After–war migration into Germany started in the mid 1950's, as a result of severe labor shortages. The early cohorts of labor migrants were actively recruited in their home countries, with housing provided upon entry, free transport, and a guaranteed work contract (see Dustmann (1996) for details). The temporary nature of migration was emphasized. Furthermore, for the type of work these early migrants were initially engaged in language was of minor importance. Immigrants were concentrated in industries which required unskilled blue collar labor. They were typically settled in especially constructed estates close to the workplace, where they lived in communities with other immigrants from the same origin country. All these factors may have contributed to lower efforts to acquire the host country language. In later years, however, and, in particular, after 1973 when active

education level and her age are found to have a positive effect on language fluency.

[6]In the US, cohort effects seem to be largely related to changes in country of origin composition. See Borjas (1987) for details.

labor recruitment came to a halt, immigrants accessed a larger range of industries. Later migrants also considered their stay as more permanent upon immigration, and both factors may have contributed to increase incentives to acquire language capital.

An alternative explanation would be that return migration is selective. Negative selective return migration would imply the opposite, however: we would then expect that workers with low language proficiency tend to return, implying that among those who remain and are included in our sample, the earlier cohorts have relatively higher levels of language proficiency than later cohorts.[7]

In all specifications, years since migration is found to have a significant positive effect on language proficiency. Comparing models 1 and 2 shows that allowing for a cohort effect increases the estimates of years since migration. [8] The average marginal effect of one additional year of residence on the probability of being fluent or very fluent, is 0.5%-points and 0.7%-points according to models 1 and 2, respectively. This is larger than the earlier finding of Dustmann (1994) for Germany, but still rather small compared to findings for other countries.[9] This suggests that the exposure effect estimated on the basis of cross–section data would be downward biased, since this

[7]To check for attrition bias, we compared estimates based upon the unbalanced panel with those for the balanced sub panel only. Results were very similar, suggesting that attrition bias is not important.

[8]If partner characteristics and the father's education level are included and the cohort effect is the only variable excluded from model 2, we find an effect of years since migration of 0.085 (standard error 0.033), i.e. still smaller than the effect in model 1.

[9]For German males, Dustmann finds an effect of 0.38%-points. For Australia, Chiswick and Miller (1995) find the effect of residence to range between 1 and 3.5 percent per year, depending on the country of origin. For Israel, Chiswick (1997) finds an effect of about 2.6 percent per year after 10 years of residence. For low skilled workers in the US, Chiswick (1991) finds that an additional year of residence increases fluency by about 3 percent.

would in fact be an estimate of the sum of the positive exposure effect (the longer the individual is in the country, the more fluent he is) and a cohort effect of the opposite sign (the older the cohort, the less fluent). Although this finding may be specific for Germany and the specific group of immigrants and time period considered, we do think it justifies the more general message that ignoring cohort effects may bias the exposure effects of years since migration.[10]

Explicitly accounting for misclassification errors improves the fit of the model, as can be seen from the difference between likelihood values of models 2, 3 and 4. In model 3, the estimated misclassification probabilities are rather precise. The probabilities of overreporting p_{12} and p_{23} are substantial, and their confidence intervals do not contain zero.[11] The probability is $p_{12} = 0.132$; it indicates that someone with bad speaking fluency has a 13 percent probability of reporting reasonable fluency in a given wave. The overreporting probability p_{21} is smaller, but still its confidence interval excludes zero. The other three misclassification probabilities are not significant. In particular, the estimates imply that individuals with good speaking fluency in German would never misclassify.

In model 4, the complete misclassification framework sketched above is used. The estimate of π_{11} is zero. This implies that there are people who sometimes overreport (67.0%) and people who sometimes underreport (14.0%), but there is no evidence of people who underreport as well as overreport. 19.0% of all people would never under-

[10]We also estimated models with nonlinear effects of years since migration, which lead to the same conclusions.

[11]Since the misclassification probabilities are by definition nonnegative, standard t-tests or likelihood ratio tests on $p_{jk} = 0$ are inappropriate (see Shapiro (1985)).

or overreport. The estimates of these group probabilities π_{rs} are not very precise, however. Although the model is identified in theory, it is hard in practice to distinguish the individual effects α_i and the idiosyncratic errors ϵ_{it} from the π_{rs} and the p_{jk}. Still, the likelihood value of model 4 is much higher than that of models 3 and 2.[12]

Some of the p_{jk} in model 4 seem quite large, suggesting that probabilities of misreporting could be substantial for the groups with a tendency to over- or underreport. To compare them with those in model 3, we should look at marginal probabilities of misclassification, taking account of the fact that we never observe in which of the three groups ((0,0), (0,1) or (1,0)) a respondent is. For example, the probability according to model 4 that someone with bad fluency reports reasonable fluency, is 0.67*0.137=0.092, compared to 0.089 in model 3. The probability that a randomly drawn individual with bad fluency in two waves, reports reasonable fluency twice, is $0.67 * 0.137^2 = 0.013$ in model 4, and $0.089^2 = 0.008$ according to model 3. The probability that someone with reasonable fluency underreports in one given wave is $0.14 * 0.444 = 0.062$ for model 4, versus 0.039 in model 3. The probability that this happens twice is $0.14 * 0.444^2 = 0.028$ for model 4, and $0.039^2 = 0.0015$ for model 3. Thus model 4 implies larger misclassification probabilities than model 3, but for the one wave probabilities, the difference is small.

The estimate of σ_ϵ reflects the importance of the idiosyncratic shocks. As expected, this is reduced in models 3 and 4 compared to models 1 and 2, in which ϵ_{it} also picks up time independent misclassification errors. Still, however, the reduction in σ_ϵ is smaller than we would have hoped; apparently, there is either more idiosyncratic noise than just misclassification errors, or our stylized model for misclassification is not able to

[12] Again, a formal chi squared test is not appropriate, due to the one-sided nature of the alternative.

pick up all misclassification errors.

The individual effects α_i are assumed to follow a distribution with four mass points. By means of normalization, one mass point is set equal to zero. We estimated models with five mass points, but the estimated probability for the fifth mass point was either very close to zero or equal to zero. The implied standard deviations of the distribution of α_i are 5.95, 5.79, 5.58, and 5.59 in models 1, 2, 3 and 4, respectively. Thus the models with explicit misclassification probabilities imply a somewhat smaller role for α_i. We would have expected, however, that the time persistent heterogeneity in terms of misclassification behavior, would reduce the role of the time persistent heterogeneity in α_i. Comparison of models 3 and 4 shows that this is not the case.

4 The Impact of Speaking Fluency on Earnings

To analyze how speaking fluency affects earnings of full-time workers, we add the following equation explaining log monthly earnings w_{it}:

$$w_{it} = x'_{it}\beta^w + \gamma y^*_{it} + \alpha^w_i + \epsilon^w_{it}\,. \tag{7}$$

We have included the underlying latent speaking fluency variable y^*_{it} instead of the discrete variable y_{it} or z_{it} because we think that y^*_{it} better reflects the impact of speaking fluency on earnings, which should not depend upon the categories that happen to have been used in the questionnaire.

As before, we assume that all errors are mean zero and we do not allow for correlation between individual effects and idiosyncratic errors, or between the error terms

and the x_{it}:

$$\epsilon_{it}^w, \alpha_i^w \text{ and } x_{it} \text{ independent}. \tag{8}$$

For the individual heterogeneity terms α_i^w, we again use a Heckman–Singer specification. We distinguish two cases:

- α_i and α_i^w independent:

$$P[(\alpha_i, \alpha_i^w) = (\alpha_l, \alpha_m^w)] = p_l p_m^w, \quad l, m = 1, \ldots, M\,. \tag{9}$$

- α_i and α_i^w not necessarily independent:

$$P[(\alpha_i, \alpha_i^w) = (\alpha_k, \alpha_k^w)] = p_k, \quad k = 1, \ldots, K\,. \tag{10}$$

According to (9), the bivariate distribution of (α_i, a_i^w) has M^2 mass points, obtained by combining the mass points of the marginal distributions. On the other hand, (10) allows for K arbitrary mass points. (9) is a special case of (10) if $K = M^2$. The results we present will be based upon $K = 3$ and $M = 9$. Comparing the results with (9) imposed with those imposing (10) will show how allowing for correlated (time persistent) unobserved heterogeneity in speaking fluency and earnings equations will affect the estimated impact of language fluency on earnings.

If the explicit misclassification errors included in the speaking fluency model are the only source of measurement error, then measurement error is automatically accounted for by including y_i^* as a right-hand variable. In this case, there seems to be no reason to allow for correlation between the idiosyncratic errors ϵ_{it} and ϵ_{it}^w. Comparing the results of the model where these misclassification errors are and are not included (models 4 and 2 in the previous section) will then show how these misclassification errors can affect the estimates of the impact of language fluency on earnings.

On the other hand, the size of the estimates of σ_ϵ in the previous section led to the conclusion that our stylized model of misclassification errors might not capture all time varying measurement error in observed speaking fluency, and that ϵ_{it} may still contain measurement error. This would mean that y_{it}^* suffers from measurement error. Following the standard argument and assuming that $\gamma > 0$, this would lead to a negative correlation between ϵ_{it}^w and ϵ_{it}. We will thus assume

$$E(\epsilon_{it}^w \epsilon_{it}) = \rho \sigma_e \sigma_e^w \tag{11}$$

and we will estimate models in which ρ is a parameter to be estimated, as well as models in which ρ is set equal to zero.

Note that the model with (9) and $\rho = 0$ implies that speaking fluency is strictly exogenous in the wage equation. In this case, we would not need exclusion restrictions on the wage equation to estimate the parameters of the wage equation. Using (10) instead of (9) (or allowing for $\rho \neq 0$) makes speaking fluency endogenous. Without exclusion restrictions, the model would then only be identified due to functional form assumptions, such as normality of the idiosyncratic error terms and the discrete distribution of the individual effects.

To identify the most general model – with correlation between errors in speaking fluency and earnings equations – nonparametrically, we need to exclude variables from the earnings equation which are in the speaking fluency equation.

For this purpose, we use dummies for the father's education level. As we have seen in the previous subsection, the father's education level has a significant impact on speaking fluency. We assume that the father's education level has no direct effect on earnings. This assumption has been criticized in the wage literature, since networking

by the father may help the child to get a better job. The immigrants in our sample, however, are first generation immigrants whose father is not in Germany, so that this argument is not valid.[13]

The regressors we include in the wage equation follow the existing literature as close as possible (see Chiswick and Miller (1995), for example). We include years of education and potential experience and its square. We also include a marital status dummy and nationality dummies. Year dummies are included to account for calendar time effects, for example reflecting rising productivity due to technical progress. Since potential experience is driven by age and education, we cannot separately identify cohort effects.

Following the existing literature in this field, we do not address potential selectivity bias due to the fact that we only use earnings of full-time workers. We thus implicitly assume that whether someone has a full-time job or not is independent of the error terms in the model, conditional upon the covariates.

Results Earnings Equations

The model is estimated with maximum likelihood, jointly with the speaking fluency equation.[14] This leads to slightly different parameter estimates for the speaking fluency equation, but the differences with the corresponding specifications in Table 2 are minor, and we therefore do not present them.

[13]We also exclude partner characteristics from the earnings equation. This hardly changes the results, however, which could be expected since the partner characteristics are insignificant in the speaking fluency equation.

[14]The Fortran code containing the likelihood function is available upon request from the authors.

In Table 4, we present the estimation results for the earnings equation, for four different specifications. In Model W1, we have used the speaking fluency equation which does not allow for explicit misclassification errors (cf. Model 2 in Table 3). Individual heterogeneity is specified through (9), not allowing for correlation between individual effects in the two equations. Idiosyncratic errors are also assumed to be independent (i.e., $\rho = 0$ in (11)). Thus this model neither corrects for correlated unobserved heterogeneity, nor for measurement errors.

We find a positive and significant effect of speaking fluency on earnings. The estimated standard deviation of y_{it}^* across individuals in this model is 9.43,[15], so the point estimate of 0.33 implies that a one standard deviation increase of y_{it}^* leads to a wage increase of about 3.1%.[16]

In Model W2, we have allowed for correlated unobserved heterogeneity, using (10) instead of (9). A likelihood ratio test suggests that this is a substantial improvement: model W1 is rejected against model W2. The estimates of model W2 imply a strong positive correlation between α_i and α_i^w: the implied estimate for the correlation coefficient is 0.56.[17]

This positive correlation implies a positive bias in the estimated effect of speaking fluency, which is removed in model W2. This explains why model W2 leads to a smaller effect of speaking fluency. The effect remains significantly positive, but an increase of

[15]We used $V_t(y_{it}^*) = V_t(x_{it}'\beta) + \sigma_\alpha^2 + \sigma_\epsilon^2$; the first of these is estimated as the sample variance of $x_{it}'\hat{\beta}$. We averaged over the seven years.

[16]Using $K = 4$ instead of $K = 3$ in (9) (as in Model 2 in Table 3) changes this to 3.1%, suggesting that this result is insensitive to the chosen number of mass points K.

[17]This is computed from the estimated distribution of (α_i, α_i^w), given in Table A2 in the appendix.

Table 4: Estimation Results, Wage Equation

	Model W1		**Model W2**		**Model W3**		**Model W4**	
	Coef	StdE	Coef	StdE	Coef	StdE	Coef	StdE
const wage	7.1100	0.0228	7.3378	0.0258	7.3943	0.0310	7.4036	0.0329
d turkish	-0.0242	0.0109	-0.0462	0.0114	-0.0344	0.0134	-0.0292	0.0138
d jugos	-0.0196	0.0115	-0.0239	0.0134	-0.0229	0.0159	-0.0478	0.0198
d greek	0.0101	0.0113	-0.0203	0.0133	-0.0176	0.0156	-0.0270	0.0165
d italian	0.0198	0.0117	-0.0027	0.0115	0.0093	0.0138	0.0023	0.0148
0.01 yrs s migr	0.0079	0.0719	0.4614	0.0951	0.4261	0.1165	0.1058	0.1735
exp	0.0322	0.0011	0.0235	0.0013	0.0244	0.0014	0.0267	0.0018
0.01 exp sq	-0.0563	0.0022	-0.0431	0.0024	-0.0446	0.0027	-0.0441	0.0027
yrs educ	0.0291	0.0013	0.0178	0.0018	0.0144	0.0022	0.0122	0.0025
married	0.1044	0.0087	0.1147	0.0093	0.1109	0.0093	0.1174	0.0099
year 85	-0.0269	0.0118	-0.0231	0.0108	-0.0226	0.0106	-0.0233	0.0106
year 86	0.0260	0.0112	0.0241	0.0101	0.0244	0.0099	0.0234	0.0099
year 87	0.0353	0.0106	0.0302	0.0095	0.0310	0.0093	0.0292	0.0094
year 89	0.1001	0.0120	0.0916	0.0113	0.0926	0.0113	0.0897	0.0114
year 91	0.1177	0.0136	0.1082	0.0124	0.1094	0.0127	0.1056	0.0131
year 93	0.1226	0.0132	0.1046	0.0125	0.1078	0.0126	0.1033	0.0130
0.01 sp fl	0.3286	0.0466	0.1363	0.0672	0.3014	0.1252	0.9122	0.2660
$\sigma(\epsilon_w)$	0.2018	0.0010	0.1873	0.0010	0.1842	0.0010	0.1865	0.0021
$\rho(\epsilon, \epsilon_w)$							-0.1537	0.0611
$\sigma(\alpha^w)^*$	0.1728		0.1927		0.1951		0.1926	
$\rho(\alpha, \alpha^w)^*$			0.5607		0.5571		0.7926	
log lik	-5071.98		-4860.38		-4788.76		-4786.39	

* $\sigma(\alpha^w)$ and $\rho(\alpha, \alpha^w)$ are computed from the estimates of the parameters in (9) (model W1) or (10) (models W2–W4), given in Table A2 in the appendix.

y_{it}^* by one standard deviation would lead to a rise in earnings of only 1.3%.

In Model W3, the misclassification probabilities are added to the speaking fluency equation (cf. Model 4 in table 3). This removes the negative bias in the speaking fluency coefficient due to time persistent and time independent misclassifications. An increase of y_{it}^* by one standard deviation (8.86 according to this model), would lead to a wage rise of 2.7%. This is only slightly smaller than the estimate in model W1; the negative and positive bias almost cancel out.

In Model W4, we have also allowed for measurement error in ϵ_{it}, i.e. we have relaxed the assumption $\rho(\epsilon,\epsilon^w) = 0$ and have estimated $\rho = \rho(\epsilon,\epsilon^w)$ in (11). The estimate of ρ is significantly negative, confirming the view that ϵ_{it} still contains measurement error, in spite of the misclassification errors that we explicitly allow for. Thus this model is a significant improvement compared to Model W3. As expected, it leads to a higher estimate of the impact of speaking fluency on earnings: an increase of y_{it}^* by one standard deviation would rise wages by about 8.1%. This is because the measurement error through ϵ_{it} is also accounted for, and thus the complete negative measurement error bias is taken out.

Most of the other coefficients vary much less across the four specifications, and generally are in line with the findings in the literature. The experience pattern is quadratic and increasing during most of the career path. Married workers earn significantly more than their unmarried colleagues. Years of education has a strong positive impact on earnings. Greek immigrants do somewhat better than immigrants from the other countries, *ceteris paribus*. Only the years since migration effect varies substantially. It is always positive, but small and insignificant in models W1 and W4, while much larger and significant in models W2 and W3.

5 Conclusions

We show that self-reported measures of speaking fluency suffer from misclassification errors. We combine a random effects ordered response model with an explicit mechanism of misclassification probabilities. We extend previous cross–section models allowing for misclassification in the literature, by distinguishing time varying and time persistent misclassification. The panel nature of our data helps to identify these two sources.

Our results indicate that the estimated probabilities of overreporting are larger than the probabilities of underreporting, some of which are virtually zero. We find some evidence of time persistent misclassification, i.e. positive correlation between misclassification events in different time periods. Neither the way misclassification is modeled, nor the assumed distribution of the individual heterogeneity, has much effect on the slope coefficient estimates in the speaking fluency equation. Thus the estimates of the determinants of speaking fluency appear to be rather robust.

An advantage compared to other, cross–section based, studies is that we can decompose the positive effect of years since migration on speaking fluency into a genuine years since migration exposure effect, and a year of entry cohort effect. We find that younger cohorts of immigrants do better than older cohorts, conditional upon years since migration and other covariates. The size of the cohort effect is smaller than the exposure effect. Not allowing for the cohort effect leads to a downward bias on the estimated exposure effect of years since migration.

We then add an earning equation for full-time workers to the model and estimate it jointly with the speaking fluency equation, allowing for misclassification and other measurement errors as well as correlated unobserved heterogeneity in earnings and

speaking fluency equations. We include the underlying continuous speaking fluency variable instead of the discrete variable which is usually included. We model unobserved heterogeneity as a discrete distribution with mass points in both equations. Our model allows us to separate the effects of measurement error and unobserved heterogeneity. We estimate various specifications and obtain a large range of positive estimates for the speaking fluency effect. Our findings suggest that correcting for measurement errors in self–reported assessments of language proficiency is crucial. Not correcting for this error leads to a substantial downward bias of the impact of speaking fluency on earnings. We also find evidence for a nonnegligible bias due to ignoring correlated unobserved heterogeneity, reflected by a positive correlation between individual effects in the two equations. In our most general model with various types of measurement errors, we find that the measurement error bias is more important than the unobserved heterogeneity bias. The estimated impact of speaking fluency according to this model is much larger than the standard models would predict.

References

- Borjas, G. J. (1985), "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants," *Journal of Labor Economics*, 3, 463-489.

- Borjas, G. J. (1987), "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 77, 531-553.

- Borjas, G. J. (1994), "The Economics of Immigration," *Journal of Economic Literature*, 32, 1667-1717.

- Butler, J.S., and R. Moffitt (1982), "A Computationally Efficient Quadrature Procedure for the One-factor Multinomial Probit Model," *Econometrica*, 50, 761-764.

- Carliner, G. (1981), "Wage differences by language group and the market for language skills in Canada," *Journal of Human Resources*, 16, 384-399.

- Chiswick, B. (1991), "Reading, speaking, and earnings among low-skilled immigrants," *Journal of Labor Economics*, 9, 149-170.

- Chiswick, B. (1997), "Hebrew Language Usage: Determinants and Effects on Earnings Among Immigrants in Israel", *Discussion Paper* 97.09, Maurice Falk Institute, Jerusalem.

- Chiswick, B. and P. Miller (1995), "The Endogeneity between Language and Earnings: International Analyses," *Journal of Labor* Economics, 13, 246-288.

- Douglas, S.M., K. Smith Conway, and G.D. Ferrier (1995), "A switching frontier model for imperfect sample separation: with an application to constrained labor supply," *International Economic Review*, 36, 503-526.

- Dustmann, C. (1994), "Speaking fluency, writing fluency and earnings of migrants," *Journal of Population Economics*, 7, 133-156.

- Dustmann, C. (1996), "Return Migration: The European Experience," *Economic Policy*, 22, 215-250.

- Grenier, G. (1984), "The effect of language characteristics on the wages of Hispanic American males," *Journal of Human Resources*, 19, 35-52.

- Hausman, J., J. Abrevaya, and F. Scott-Morton (1997), "Misclassification of the dependent variable in a discrete response setting," *Journal of Econometrics*, forthcoming.

- Heckman, J. J. and B. Singer (1984), "A method for minimizing the impact of distributional assumptions in econometric models for duration data," *Econometrica*, 52, 271-320.

- Kossoudji, S. (1988), "English language ability and the labor market opportunities of Hispanic and East Asian immigrant men," *Journal of Labor Economics*, 6, 202-228.

- Lee, L.-F, and R.H. Porter (1984), "Switching regression models with imperfect sample separation information: With an application on Cartel stability," *Econometrica*, 52, 391-418.

- McManus, W., W. Gould, and F. Welch (1983), "Earnings of Hispanic men: the role of English language proficiency," *Journal of Labor Economics*, 1, 101-130.

- Rivera-Batiz, F. (1990), "English language proficiency and the economic progress of immigrants," *Economics Letters* , 34, 295-300.

- Shapiro, A. (1985), "Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints," *Biometrica* , 72, 133-144.

- Wagner, G., R.V. Burkhauser, and F. Behringer (1993), "The English Public Use File of the German Socio Economic Panel," *Journal of Human Resources*, 28, 429-433.

Appendix

Table A1: Description and Summary Statistics for Wave 1

Code	Mean	StD	Explanation
yrs s migr	17.722	5.817	Years of Residence in Germany
age	41.234	10.769	Age
age entry	23.512	8.648	Age at Entry
d turkish	0.317	0.465	Dummy; 1 if Turkish
d jugos	0.197	0.398	Dummy; 1 if Yugoslavian
d greek	0.139	0.346	Dummy; 1 if Greek
d italian	0.208	0.406	Dummy; 1 if Italian
d spanish	0.136	0.343	Dummy; 1 if Spanish
yrs edu	9.941	2.042	Years of Schooling
mar, p G	0.056	0.230	Dummy; 1 if Married, Partner German
mar	0.787	0.408	Dummy; 1 if Married, Partner not German
p yrs edu	9.125	1.983	Years of Schooling, Partner
p yrs s migr	14.67	7.144	Years of Residence, Partner
p age	39.72	9.386	Age, Partner
n children	1.206	1.248	Number of children
f educ l 1	0.226	0.418	Father no education
f educ l 2	0.332	0.471	Father primary education
f educ l 3	0.292	0.455	Father basic education
f educ l 4	0.038	0.191	Father intermediate education
f educ l 5	0.005	0.073	Father secondary education
f educ l 6	0.104	0.305	Father education missing

Table A2: Distribution of (α_i, α_i^w) in models W1–W4

	Model W1		Model W2		Model W3		Model W4	
	Coef	StdE	Coef	StdE	Coef	StdE	Coef	StdE
p_1	0.2828	0.0195						
p_2	0.5564	0.0210						
α_1	16.8339	0.3982						
α_2	8.3368	0.2472						
p_1^w	-0.1995	0.0152						
p_2^w	0.1261	0.0125						
α_1^w	-0.3011	0.0064						
α_2^w	0.3112	0.0075						
p_1			0.0952	0.0129	0.0677	0.0122	0.0665	0.0122
p_2			-0.0080	0.0029	-0.0085	0.0030	-0.0080	0.0029
p_3			0.1346	0.0154	0.0994	0.0154	0.0945	0.0151
p_4			0.0259	0.0058	0.0255	0.0058	0.0254	0.0059
p_5			0.0231	0.0055	0.0271	0.0066	0.0276	0.0065
p_6			0.3578	0.0205	0.3547	0.0254	0.3596	0.0252
p_7			0.1131	0.0129	0.1232	0.0142	0.1233	0.0142
p_8			-0.1291	0.0144	-0.1372	0.0167	-0.1430	0.0170
α_1			15.6928	0.5192	11.1143	0.7252	11.3479	0.7464
α_2			17.0529	2.5980	12.7387	2.7276	13.0140	2.8486
α_3			18.6515	0.6119	15.5085	0.8787	16.0864	0.9293
α_4			9.7861	0.5836	6.1849	0.7489	6.3995	0.7661
α_5			20.0141	1.2440	8.9212	0.7297	9.2649	0.7419
α_6			9.3325	0.3247	5.7677	0.4538	5.9544	0.4577
α_7			4.2243	0.3749	-1.0288	0.6182	-0.8580	0.6278
α_8			10.0403	0.3804	8.1679	0.5609	8.5507	0.5698
α_1^w			-0.2665	0.0169	-0.3475	0.0232	-0.4209	0.0353
α_2^w			0.7746	0.0408	0.7255	0.0383	0.6502	0.0506
α_3^w			-0.0069	0.0167	-0.0998	0.0262	-0.2096	0.0461
α_4^w			-0.6809	0.0163	-0.7293	0.0184	-0.7692	0.0235
α_5^w			0.3789	0.0272	0.3471	0.0328	0.2892	0.0403
α_6^w			-0.0772	0.0112	-0.1375	0.0135	-0.1756	0.0191
α_7^w			-0.2822	0.0124	-0.3047	0.0133	-0.3024	0.0148
α_8^w			0.1656	0.0134	0.0984	0.0192	0.0390	0.0278

Note: Parameters are defined in (9) (model W1) or (10) (models W2–W4).

See Table 4 for parameters of interest.